Serving NASA HDF-EOS Data through NWGISS Coverage Server

Wenli Yang and Liping Di
Laboratory for Advanced Information Technologies and Standards
Center for Earth Observing and Space Research
School of Computational Sciences
George Mason University
9801 Greenbelt Road, Suite 316-317
Lanham, MD 20706

Abstract - This paper describes the Web Coverage Server component of the NASA Web GIS software suite (NWGISS) that provides HDF-EOS data through the web using OpenGIS Consortium (OGC) Web Coverage Service (WCS) specification. The NWGISS Web Coverage Server (NWGISS WCS) is designed to allow clients of geographic information systems (GIS) to access multi-dimensional, multi-temporal HDF-EOS data. It has been implemented with three interface protocols: GetCapabilities, DescribeCoverageType, and GetCoverage. In addition to HDF-EOS, the server can encode the returned data in several widely used formats, including binary, geoTiff, NITF, and GIF. The server provides on-the-fly georectification function that can extract swath data and georectified it into user requested spatial resolutions. This functionality is important to many HDF-EOS data users.

I. Introduction

The Hierarchical Data Format (HDF) is a widely used scientific data format because of its portability and multiple data model support. NASA's Earth Observing System (EOS) project extended HDF to HDF-EOS by adding three new EOS specific data models – point, swath, and grid [1]. HDF-EOS is the standard format for EOS data and products. EOS project has since produced tremendous amount of data in HDF-EOS format and these data are highly demanded by a broad range of research and application communities. GIS is one of the important tools for analyzing NASA's EOS data. However, current GIS systems use different internal formats and are very difficult to interoperate with each other and with EOS data information system (EOSDIS) directly. Most GIS software packages are also incapable of ingesting data in HDF-EOS. Therefore, the development of capability for delivering GIS-ready EOS data directly to user's GIS systems through Internet for analysis will greatly enhance the interoperability and public use of EOS data.

The objective of NWGISS WCS is to make HDF-EOS data available to GIS applications according to OGC Web GIS Specifications. Through the development of NWGISS WCS, we will evaluate the suitability of OGC interface standards for making NASA data accessible to GIS community. The NWGISS WCS works with the two HDF-EOS grided data models, i.e., swath and grid, and is easy-to-use and easy-to-setup. The development of NWGISS WCS will make NASA EOS data immediately available and accessible to GIS developed by major GIS vendors. Hence, it will significantly increase the accessibility, interoperability, and inter-use of HDF-EOS data, improve the data analysis and visualization, promote the use of HDF-EOS data not only in the global change research but also in the public concerned issues such as environment and resource management, education, and community planning.

II. The OGC Web mapping testbed and web services initiative

The Open GIS Consortium, Inc. (OGC) is a not-for-profit membership organization founded in 1994 to address the lack of interoperability among systems that process geospatial data, and between these systems and mainstream computing systems. OGC's mission is to give the world's information systems a new connection to physical reality by making georeferenced data behave like just another standard data type in systems of all kinds. To achieve that mission, OGC has successfully engaged key user organizations and technology providers in a consensus process to develop technology standards and business process innovations that support widespread adoption and use of georeferenced data and services.

OGC had successfully implemented two Web Mapping Testbeds, WMT I and WMT II [2]. The testbeds produced a set of web-based interoperability specifications. WMT I resulted

in an OGC Web Mapping Specification (WMS) version 1. WMS allows interactively assembling maps from multiple servers. WMT II was finished in December 2000 and produced a set of new interoperability specifications. One of the most important specifications for NASA from WMT II is the Web Coverage Specification (WCS). It allows a WCS client to access real multi-dimensional, multi-temporal data from coverage servers. WCS provides an interoperable way of accessing geospatial data, especially those from remote sensing.

OGC Web Services (OWS) represent an evolutionary, standards-based framework that enable seamless integration of a variety of online geoprocessing and location services [3]. OWS was launched in 2001 and the first phase ended with a demo in April, 2002. NWGISS WCS was an important part of the OWS demo and was the only WCS that can serve NASA HDF-EOS products. OWS phase II is just started and a demo will be held in November of 2002.

III. THE OGC WEB COVERAGE SERVICE SPECIFICATION

The OGC Web Coverage Service Specification is designed for enabling GIS clients to access real multi-dimensional, multi-temporal data. WCS specification defines two required interface protocols, namely GetCapabilities and GetCoverage. A third optional protocol, DescribeCoverageLayer was added in versions 0.5, 0.6, and 0.7 [4].

The GetCapabilities interface is designed to let OGC web coverage clients obtain information about the server capabilities as well as the availability of coverage layers. With these pieces of information, a client can formulate a GetCoverage request based on user requirements. The capability description is encoded in XML. The OGC specification provides the XML schema for the description.

The GetCoverage interface is designed for clients to request multidimensional data by specifying extents in the spatial, temporal, elevation, band, and other dimensions, and the encoding format for the returning coverage. The server has to extract the data from archives based on client's specifications, package the data in the format specified by the client, and return the data to client.

The optional DescribeCoverageLayer interface lets clients request a full description of any coverage layers served by a the server. It is different from the getCapabilities interface in that DescribeCoverageLayer gives information about a specific data coverage layer while getCapabilities gives overall description of coverage layers and service. In the latest version of WCS specification (version 0.7), the Capabilities XML contains detailed descriptions of all available coverage layers. Thus, a DescribeCoverageLayer request will result in one or more subsets from the Capabilities XML.

IV. THE NWGISS WEB COVERAGE SERVER

NWGISS WCS is implemented according to the OGC WCS specification versions 0.5 and 0.6 and is ready to be upgraded to include version 0.7, which was first released for discussion on April 04, 2000, as soon as version 0.7 is stabilized.

A. Data format supported by NWGISS WCS

NWGISS WCS is specifically designed for NASA EOS data products. Thus, only NASA HDF-EOS formatted data can be served in this server. We have learned that that a few data producers use native HDF rather than HDF-EOS in generating satellite remote sensing data. Because native HDF does not provide a standard method to store geolocation information, it is not possible to serve native HDF file in a WCS. We have developed a native HDF to HDF-EOS conversion tool for certain HDF products so that those data can be made available to web coverage clients through WCS.

B. Returning Data Format from NWGISS WCS

OGC WCS specification does not prescribe any particular encoding for coverage replies. NWGISS WCS provides a few widely used data encoding. These include binary, HDF-EOS, NITF [5], and GeoTIFF [6]. In the binary encoding, science data sets and geolocation data are written in different plain binary files and information about data sets, including number and sizes and dimension, data types, geographic projections or geo/data mapping relationships, is written in a separate ASCII text file. All these separate files are zipped into a single file and returned to the client. In the HDF-EOS encoding, the original metadata is copied into the returned file, with modifications to the spatial

bounding box information if it exists in the original HDF-EOS file. The original data structures (e.g., swath, grid, names of swaths, grids, fields, dimensions) are kept unchanged. NITF can only handle georectified data (i.e., NASA level 3 data). It is not very useful for NASA levels 1 and 2 data because it is not capable of handle geolocation information in swath data. GeoTiff format provides limited capability of storing geo/data mapping information through geo-tie points or transform matrix but is not very suitable for most NASA levels 1 and 2 data. In addition to the aforementioned four encoding methods, NWGISS WCS also provide GIF encoding upon returning. Since GIF is an image format, this encoding does not provide original coverage data, but the image(s) derived from the coverage data. It is essentially a Web Mapping Service functionality and is provided here as a value-add component, which will be useful in future development of a Web Coverage Portrayal Server.

C. Subsetting in NWGISS WCS

NWGISS WCS supports multidimensional subsetting, both spatial and non-spatial. Spatial subsetting is based on spatial bounding box provided by the clients and non-spatial subsetting is based on dimension name and a subsetting value, such as Band_1KM_Emissive=1/16/3. For georectified data (i.e., HDF-EOS grid), the returned coverage exactly matches the requested spatial bounding box. For ungeorectified data (i.e., HDF-EOS swath), the returned coverage represents the minimum subset to fill the spatial bounding box sent by the client. That is, no single row (scan) or column (cross-scan) can be drop from the returned coverage. Otherwise the requested bounding box will not be filled. This is a significant improvement to some existing software packages that retrieve a complete swath scan once the scan touches the bounding box.

D. On-the-fly Georectification in NWGISS WCS

On-the-fly georectification is designed for getting georectified coverage from HDF-EOS swath data. A further modification to the minimum coverage subsetting on swath data is to georectify the swath subset into grid so that it can be fitted exactly into the requested bounding box. Two georectification algorithms are implemented in NWGISS WCS: a bivariate polynomial regression approach and a piecewise bilinear interpolation approach.

The bivariate polynomial regression approach is effective and accurate when geometric distortions can be modeled by low order polynomials (usually two- to four-order) and adequate number of ground control points (GCP) is available. For data with significant geometric distortions, such as airborne images, global polynomial fitting usually cannot generate satisfactory results. A remedy to this problem is using piecewise rectification approaches in which polynomial regressions are performed locally and transform piece by piece the original image to map/earth coordinate using locally derived fitting functions [7][8]. We implemented the global bivariate polynomial regression algorithm in NWGISS WCS. The algorithm produced satisfactory results for data with relatively small distortion or within a small bounding box (e.g., ASTER data or 1/4 to 1/6 scan/frame of MODIS L1B data) but significant errors were observed for data having wide scan angles (e.g., MODIS L1B full granule data).

Many satellite data products contain very dense geolocated pixels, such as the level 1B AVHRR data and MODIS data. For those data, the piecewise bilinear interpolation algorithm is more appropriate. The algorithm utilizes every available geolocation pixel and produces more accurate rectified images in many cases. The algorithm was first developed for processing AVHRR level 1B data stored in half inch, 9-track tapes [9]. It is modified and implemented in the NWGISS WCS based on HDF-EOS swath structure. The algorithm involves a spatial interpolation procedure that transfers pixels of a data field from raw image coordinate to map/earth coordinate, and a data value interpolation that fills blank pixels in the map/earth coordinate. This algorithm performs extremely accurate in our testing with MODIS and ASTER data. The position error is less than one-tenth of a pixel (assuming the geolocation data in swath are error free).

E. Multiple Spatial Reference System Support in NWGISS WCS

The OGC WCS specification versions 0.5 to 0.7 provide two ways in defining spatial bounding box: a) a single spatial reference system (SRS) for both request and returning data; and b) one SRS for request and another for returning data.

With the on-the-fly georectification capability, NWGISS WCS can easily handle the multi-SRS requests. For example, a client may specify bounding box in row/column (scan/frame) coordinate and request returning data in latitude/longitude (lat/lon) coordinate, or vise versa. As long as the returning SRS is not in row/column coordinate, on-the-fly georectification is invoked for swath coverage.

F. NWGISS WCS websites and test queries

NWGISS WCS can be accessed at the following address:
http://laits.gmu.edu/cgi-bin/NWGISS/wcshdfeos
http://heineken.gsfc.nasa.gov:8080/cgi-bin/dib/wcshdfeos

The followings are sample queries that request capabilities or HDF-EOS data from NWGISS WCS:

http://laits.gmu.edu/cgi-bin/NWGISS/wcshdfeos?service=wcs&REQUEST=capabilities
http://laits.gmu.edu/cgi-bin/NWGISS/wcshdfeos?VERSION=0.5&REQUEST=COVERAGE&BBOX=-120,22,-110,30&FORMAT=HDFEOS&LAYERS=mod2.hdf:SWATH:MODIS_SWATH_Type_L1B:EV_1KM_Emissive&SRS=EPSG:4326&Band_1KM_Emissive=1/16/1&exception=xml
http://laits.gmu.edu/cgi-bin/NWGISS/wcshdfeos?service=wcs&request=coverage&bbox=20.65,1.5,96.85,25&layers=india_2kb1_rect.hdf:GRID:AVHRR:Band1&srs=EPSG:4326&format=geoTiff&exception=xml

V. FUTURE WORK

The current NWGISS WCS is fully functional with all current OGC WCS specification versions (up to version 0.6). Further upgrade will be needed as new versions emerge, e.g., version 0.7, which is under discussion and modification. With the start of OWS phase II, more upgrade to the server is expected. In addition, we believe that coordinate reprojection functionality should be implemented. Currently, consensus has not been reached as if the reprojection should be included in WCS or should be separate service. We believe that a minimal reprojection capability is necessary as part of WCS. Currently, we have implemented a bounding box reprojection function between integerized sinusoidal (ISIN) projection and lat/lon coordinate. This is essential because a large number of EOS data are in ISIN projection and most users will request data in lat/lon rather than ISIN bounding box. Thus, an ISIN to lat/lon reprojection for the bounding box corners is necessary. Reprojections between other SRSs are also needed. Other expected developments include implementing full functional DescribeCoverageLayer capability, parsing XML encoded post request (currently only implemented for version 0.6 specification), encoding for other popular remote sensing and GIS formats such as ARC/INFO grid and shapefile, building multi-temporal database and improving/adding temporal subsetting capability. This last capability is especially worth to be mentioned. The multi-temporal subsetting is limited to dimension subsetting. That is, time is a dimension in multidimensional data (such as TOMS ozone data that contain temporal dimension). Multi-temporal information is also presented in multiple file's metadata, such as multiple MODIS granule with time stamp included in metadata. Because there is no standard as to how to store time information in metadata for different HDF-EOS files, we have not implemented the temporal search in metadata. We expect to start from product specific capability first and gradually to include most, if not all, EOS products.

ACKNOWLEDGEMENT

This project was supported mainly by grants from NASA Earth Science Technology Office (ESTO). Additional funding was provided by Open GIS Consortium (OGC) for the development of NWGISS coverage server as a part of OGC WMT II.

REFERENCES

[1] NASA, HDF-EOS Library User's Guide for the ECS Project, NASA Technical Peper, 170-TP-500-001, 1999.
[2]OGC,http://www.opengis.org/pressrm/summaries/20011107.TS.WMT2.htm
[3] OGC, Introduction to OGC Web Services, A. Doyle and C. Reed, eds, 2001, http://ip.opengis.org/ows/index.html
[4] OGC, Web Coverage Sevice (WCS), versions 0.5, 0.6, 0.7, J. Evans, eds, 2000, 2001, 2002, http://www.opengis.org/
[5] NIMA, National Imagery Transmission Format Version 2.1, MIL-STD-2500B, 1997.
[6] GeoTIFF Working Group, GeoTIFF format Specification, Revision 1.0, 2000, http://www.remotesensing.org/geotiff/spec/geotiffhome.html
[7] M. Ji and J. R. Jensen, "Continuous piecewise geometric rectification for airborne multispectral scanner imagery",

Photogrammetric Engineering and Remote Sensing, vol. 66, pp.163-171, 2000.

[8] E. J. Christensen, J. R. Jensen, E. W. Ramsey, H. E. Mackley, Jr., "Aircraft MSS data registration and vegetation classification for wetland change detection", *Photogrammetric Engineering and Remote Sensing*, vol.53, pp.521-529, 1988.

[9] Di, L., and Rundquist D., 1994. "A One-Step Algorithm for Geometric and Radiometric Correction and Calibration of AVHRR level 1b DATA," *Photogrammetric Engineering and Remote Sensing.* Vol. 60, No. 2, pp. 165-171.